Some Results on Symmetric DGSMS and DGSM Equivalence *

David Haussler

CU-CS-199-79

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

* This research was conducted while the author was supported by a University of Colorado doctoral fellowship. The author gratefully acknowledges financial support of NSF grant number MCS 79-03838.

SOME RESULTS ON SYMMETRIC DGSMS AND DGSM EQUIVALENCE

by

David Haussler
Department of Computer Science
University of Colorado at Boulder
Boulder, Colorado

CU-CS-199-79 November 1979
Revised February 1981

This research was conducted while the author was supported by a University of Colorado doctoral fellowship.

Abstract

A new operation on languages is introduced which is related to the complete twin shuffle. Symmetric DGSM's are characterized in terms of this operation and it is shown that:

1. It is decidable whether or not two DGSM mappings are equivalent on a regular set.
2. It is decidable whether or not an augmented DGSM mapping is symmetric.

Let us begin by defining the mapping REDUCT, which plays a central role in our results.

<u>Definition</u>: The mapping $\mathrm{REDUCT}_\Sigma : (\Sigma \cup \overline{\Sigma})^* \to (\Sigma \cup \overline{\Sigma})^*$ is defined as the input-output mapping of the following one-way two-headed transducer:

1. Given a word $w = a_1 \ldots a_n \in (\Sigma \cup \overline{\Sigma})^*$, head 1 is placed over the first unbarred symbol and head 2 is placed over the first barred symbol, reading from left to right. Let us call the symbol under head 1 $S(H_1)$ and the symbol under head 2 $S(H_2)$.

2. From this starting position proceed as follows: while both heads are still on w and $\overline{S(H_1)} = S(H_2)$, repeat: erase $S(H_1)$ and $S(H_2)$ and then move H_1 to the next unbarred symbol to its right and H_2 to the next barred symbol to its right.

3. What is left when no further erasures or moves can be made is $\mathrm{REDUCT}_\Sigma(w)$.

When the alphabet Σ is understood or irrelevant, we will abbreviate REDUCT_Σ as REDUCT. It will be convenient to define a mapping TƆUᗡƎЯ in a similar fashion, except moving from right to left. Both mappings are extended to languages: for $L \subseteq (\Sigma \cup \overline{\Sigma})^*$, $\mathrm{REDUCT}(L) = \{\mathrm{REDUCT}(w) : w \in L\}$. It should be noted that

$$\mathrm{REDUCT}_\Sigma^{-1}(\lambda) = \text{TƆUᗡƎЯ}_\Sigma^{-1}(\lambda) = L_\Sigma,$$

where L_Σ is the complete twin shuffle over Σ defined in [2].

The operation REDUCT is a rather powerful one as is indicated by the following result: for any recursively enumerable set $K \subset \Sigma^*$ there exists an alphabet $\Delta \supset \Sigma$ and a regular set $R \subset (\Delta \cup \overline{\Delta})^*$ such that

$$K = \mathrm{REDUCT}_\Delta(R) \cap \Sigma^*.$$

For the proof of this fact, the reader is referred to Theorem IV of [3] and comments following. As a result REDUCT(R) is in general not recursive for regular R. However, we will show that it is decidable whether or not REDUCT(R) is finite for regular R.

We begin with a few definitions and lemmas.

Definition: Let $A = \langle Q, \Sigma, \delta, q_{in}, F\rangle$ be a finite automaton which accepts a nonempty set $R \subseteq \Sigma^*$. For each ordered pair $\langle p,q\rangle \in Q \times Q$ we define the set $D_{pq} = \{w = a_1 \ldots a_n : n \geq 1,\ a_i \in \Sigma$ for $1 \leq i \leq n,\ \delta(p,w) = q$ and, for all $1 \leq i < j \leq n,\ \delta(p,a_1 \ldots a_i) \neq \delta(p,a_1 \ldots a_j)\}$.

The core of A, denoted C_A, is defined by

$$C_A = \bigcup_{q \in F} D_{q_{in}q}.$$

The set of accessed states of A, denoted Q_0, is defined by $Q_0 = \{q :$ there exists a $w \in R$, such that $w = w_1 w_2$ and $\delta(q_{in}, w_1) = q\}$.

Definition: Let Δ and Σ be alphabets such that $\Sigma \subseteq \Delta$. Then $\mathrm{PRES} : \Delta^* \to \Sigma^*$ is defined as

$$\mathrm{PRES}_\Sigma(a) = a \quad \text{for } a \in \Sigma$$

$$\mathrm{PRES}_\Sigma(a) = \lambda \text{ otherwise.}$$

First we have a combinatorial lemma.

Lemma I: Given $x,y,z \in \Sigma^*$ with $x,y \neq \lambda$, let $d = \mathrm{GCD}(|x|,|y|)$ and $k_1 = |y|/d$, $k_2 = |x|/d$. Then $z x^\omega = y^\omega$ if and only if there exists a nonnegative integer k and words $\alpha, \beta \in \Sigma^*$ such that $x^{k_1} = \alpha\beta$, $y^{k_2} = \beta\alpha$ and $z = \beta(\alpha\beta)^k$.

Proof:

The "if" part is obvious.

For the "only if" part, we observe that since $|x^{k_1}| = |y^{k_2}|$, if $z x^\omega = y^\omega$ then y^{k_2} is a prefix of $z x^{k_1}$ so that $z x^{k_1} = y^{k_2} w_0$ for some w_0 with $|w_0| = |z|$. But then since $z x^\omega = y^\omega$, w_0 is a prefix of y^ω. Thus w_0 equals z and we have $z x^{k_1} = y^{k_2} z$. Now if $|z| \leq |x^{k_1}|$, then $\exists w : z x^{k_1} = z w z = y^{k_2} z$ which implies $x^{k_1} = w z$ and $y^{k_2} = z w$. Let $k = 0$, $z = \beta$, $w = \alpha$. We have our result. Otherwise if $|z| > |x^{k_1}|$, then $\exists z_0 : z x^{k_1} = y^{k_2} z_0 x^{k_1} = y^{k_2} z$ which implies that $z = z_0 x^{k_1}$ and thus $z_0 x^\omega = y^\omega$. Since $|z_0| < |z|$, a simple inductive argument establishes our claim $\square$

Now we have our main lemma.

Lemma II: Let $A = \langle 0, \Sigma \cup \bar{\Sigma}, \delta, q_{in}, F \rangle$ accept $R \subset (\Sigma \cup \bar{\Sigma})^*$

Then the following are equivalent:

(i) REDUCT(R) is finite.

(ii) For all $q \in Q_0$, $w_1 \in D_{q_{in}q}$ and $w_2 \in D_{qq}$

1. $\overline{|\text{PRES}_\Sigma(w_2)|} = \overline{|\text{PRES}_{\bar{\Sigma}}(w_2)|}$

and 2. $\text{PRES}_\Sigma(w_1)\ (\text{PRES}_\Sigma(w_2))^\omega = \text{PRES}_{\bar{\Sigma}}(w_1)\ (\text{PRES}_{\bar{\Sigma}}(w_2))^\omega$.

(iii) $\text{REDUCT}(R) = \text{REDUCT}(C_A)$.

Proof:

(i) → (ii)

Assume $\exists\, q \in Q_0$, $w_1 \in D_{q_{in}q}$ and $w_2 \in D_{qq}$ such that ii.2 fails. Then $\exists$ K such that our reduction machine stops before either head has reached the end of $w_1 w_2^K$. Let $x = \text{REDUCT}(w_1 w_2^K)$. Since $q \in Q_0$, $\exists\, w_3 : \forall n\; w_1 w_2^n w_3 \in R$. But then $\text{REDUCT}(\{w_1 w_2^n w_3 : n \in \mathbb{N}\}) = \{x w_2^n w_3 : n \in \mathbb{N}\}$ which is infinite. Thus REDUCT(R) is infinite. Assume $\exists\, q \in Q_0$, $w_2 \in D_{qq}$ such that ii.1 fails. If $K = \left\| |\text{PRES}_\Sigma(w_2)| - |\text{PRES}_{\overline{\Sigma}}(w_2)| \right\|$ then $K > 0$. Since $q \in Q_0, \exists\, w_1 \in D_{q_{in}q}$ and w_3 such that $\forall n\; w_1 w_2^n w_3 \in R$. But since for all n $|w_1 w_2^n w_3| \geq |\text{REDUCT}(w_1 w_2^n w_3)| \geq nk - |w_1| - |w_3|$ again $\text{REDUCT}(\{w_1 w_2^n w_3 : n \in \mathbb{N}\})$ is infinite.

(ii) → (iii)

If ii.2 succeeds then either $\text{PRES}_{\overline{\Sigma}}(w_1)$ is a prefix of $\overline{\text{PRES}_\Sigma(w_1)}$ or vice versa. Assume the former, without loss of generality. Then $\text{REDUCT}(w_1) \in \Sigma^*$. Let $z = \text{REDUCT}(w_1)$, $x = \text{PRES}_\Sigma(w_2)$ and $y = \text{PRES}_{\overline{\Sigma}}(w_2)$. ii.2 then reduces to $\overline{z}\,\overline{x}^{\omega} = y^{\omega}$. If ii.1 succeeds then $|x| = |y|$ and by Lemma I, $\exists\, \alpha, \beta : x = \alpha\beta$, $\overline{y} = \overline{\beta}\,\overline{\alpha}$ and $z = \beta(a\beta)^K$ for some K. We have then: $\text{REDUCT}(w_1 w_2) = \text{REDUCT}(\text{REDUCT}(w_1) w_2) = \text{REDUCT}(\beta(\alpha\beta)^K w_2) = (\beta\alpha)^K\beta = z = \text{REDUCT}(w_1)$ as a simple computation verifies.

We demonstrate now that $\text{REDUCT}(R) = \text{REDUCT}(C_A)$. For any word $w \in R - C_A$, let q be the first state which occurs twice while A is accepting w. Thus $w = w_1 w_2 w_3$ where $w_1 \in D_{q_{in}q}$ and $w_2 \in D_{qq}$ for some w_1, w_2, w_3. We have then $\text{REDUCT}(w_1 w_2 w_3) = \text{REDUCT}(\text{REDUCT}(w_1 w_2) w_3) = \text{REDUCT}(\text{REDUCT}(w_1) w_3) = \text{REDUCT}(w_1 w_3)$. Continuing in this manner, we will find $w_0 \in C_A$ such that $\text{REDUCT}(w) = \text{REDUCT}(w_0)$. Thus $\text{REDUCT}(R) = \text{REDUCT}(C_A)$.

(iii) ⇒ (i)

Obvious since C_A is finite.

□

Theorem I: Let Σ be a finite alphabet and let $R \subseteq (\Sigma \cup \bar{\Sigma})^*$ be a regular set. Then it is decidable whether or not $REDUCT_{\Sigma}(R)$ is finite. □

Proof:

The result follows directly from Lemma II. Given A accepting R, we need only check whether or not the conditions of part (ii) of Lemma II are satisfied for each triple $<q,w_1,w_2>$, where $q \in Q_o$, $w_1 \in D_{q_{in}q}$ and $w_2 \in D_{qq}$. That this can be done effectively for each of the finitely many such triples is given by Lemma I. □

Definition: We now introduce the cross product of two DGSM's.

If $A_1 = < Q_1, \Sigma, \Delta, \delta_1 q_{in_1}, F_1 >$

and $A_2 = < Q_2, \Sigma, \Delta, \delta_2, q_{in_2}, F_2 >$

then $A_1 \times A_2 = < Q_1 \times Q_2, \Sigma, \Delta, \delta, < q_{in_1}, q_{in_2} >, F_1 \times F_2 >$

where

$$\delta(< q_i, q_j >, a) = (< \delta_{1s}(q_i, a), \delta_{2s}(q_j, a) >, \delta_{1o}(q_i, a) \overline{\delta_{2o}(q_j, a)})$$

where δ_s gives the new state, δ_o the output.

Definition: The output of DGSM A on w will be denoted A(w), similarly for any language, $A(L) = \{A(w) : w \in L\}$. The set accepted by A we denote DOM(A). A(DOM(A)) will be called TR(A) for translation, following [1]. For simplicity, we assume that A(w) is defined whether or not w is accepted.

Theorem II: It is decidable whether or not two DGSMs A_1 and A_2 are equivalent on a regular set $R \subseteq \Sigma^*$.

Proof:

A_1 and A_2 are equivalent on R iff $REDUCT_{\Sigma}((A_1 \times A_2)(R)) = \{\lambda\}$. Since $R' = (A_1 \times A_2)(R)$ is regular, we can use Lemma II to check if $REDUCT_{\Sigma}(R')$ is finite. If so we can go on to check if $REDUCT_{\Sigma}(R') = \{\lambda\}$. □

For our next theorem, recall from [1] that a DGSM mapping M is symmetric iff there is a reversed DGSM A^R which reads from right to left, has the same domain as M, and such that $M(w) = A^R(w)$ for all $w \in DOM(M)$. Given a DGSM A realizing M, A^R is called a symmetric partner for A and $< A, A^R >$ is called a symmetric pair. A prefix bound on $< A, A^R >$ is a number s such that:

$$\forall\ w = v_1 v_2 \in DOM(A), \left\| |A(v_1)| - (|A^R(v_1 v_2)| - |A^R(v_2)|) \right\| < s.$$

It was shown in [1] that such a bound exists for each symmetric pair. Also in [1] we have that given a mapping $M : \Sigma^* \to \Delta^*$, $AUG(M) : \$\Sigma^*\$ \to \$\Delta^*\$$ is defined as $AUG(M)(\$w\$) = \$A(w)\$$ where $\$ \notin \Sigma \cup \Delta$.

Any DGSM mapping $M : \$\Sigma^*\$ \to \$\Delta^*\$$ where $\$ \notin \Sigma \cup \Delta$ is considered here to be an augmented DGSM mapping, even if the corresponding mapping from $\Sigma^* \to \Delta^*$ is not a DGSM mapping, as in the mapping defined:

$$M(\$w\$) = \begin{cases} \$w\,e\$ & \text{if } |w| \text{ is even} \\ \$w\,o\$ & \text{if } |w| \text{ is odd} \end{cases}$$

Augmenting essentially provides endmarkers to let machines reading in either direction know when they are reading the final letter of the input. We consider augmented DGSM mappings because they constitute a natural extension of the important class of DIL mappings (see [1]) and are themselves special cases of mappings induced by sweeping and two way automata with output (see [4]). As was pointed out in [1], the augmented symmetric mappings form an interesting class of languages between the classes of DIL mappings and augmented DGSM mappings.

We now give a characterization of augmented symmetric DGSM mappings using REDUCT. First we have two definitions and two simple lemmas.

<u>Definition</u>: For any DGSM A with state q, let A_q be A started in state q.

<u>Definition</u>: Two DGSM mappings M_1 and M_2 are defined to be almost equivalent from the right iff there exists an integer K such that for **any**

$w \in DOM(M_1) \cap DOM(M_2)$ there exist words u, v, x such that

$$M_1(w) = u\,x$$

$$M_2(w) = v\,x$$

with $|u|, |v| \leq K$.

The least such K, if it exists, is called the prefix divergence bound for M_1 and M_2.

Lemma III: Let A_1 and A_2 be DGSM's realizing mappings M_1 and M_2 respectively. Then M_1 and M_2 are almost equivalent from the right iff TƆUDƎЯ($TR(A_1 x A_2)$) is finite.

Proof:

For the "if" part we may take K larger than the length of the longest word in TƆUDƎЯ($TR(A_1 x A_2)$).

For the "only if" part we note that card(TƆUDƎЯ($TR(A_1 x A_2)$)) $< |\Sigma|^{2K}$ if $|\Sigma| > 1$ (otherwise it is less than $K+1$). □

Lemma IV: Let $M_1, \ldots, M_n$ be DGSM mappings such that any two mappings M_i and M_j $1 \leq i, j \leq n$ are almost equivalent from the right. Then there exists a K such that for any $w \in \bigcap_{i=1}^{n} DOM(M_i)$ there exist words $\{u_1, \ldots, u_n\}$ and x such that $\forall\, i, 1 \leq i \leq n, M_i(w) = u_i x$ and $|u_i| < K$.

Proof:

Let K be the largest prefix divergence bound for any of the pairs (M_i, M_j). For any fixed w find i such that $|M_i(w)|$ is maximal among the $M_j(w)$'s. Then either $|M_i(w)| < K$, in which case the result trivially follows, or all the $M_j(w)$'s share a common suffix with $M_i(w)$ of length $\geq |M_i(w)| - K + 1$, from which the result also follows, since $|M_i(w)|$ was maximal. □

We may call K the prefix divergence bound for the set of mappings $\{M_1, \ldots, M_n\}$.

We are now ready to state and prove our theorem.

<u>Theorem III</u>: An augmented DGSM mapping M is symmetric iff for any DGSM A realizing M and any states q_i, q_j of A, reachable from the initial state. $REDUCT_R(TR(A_{q_i} \times A_{q_j}))$ is finite, i.e., the mappings induced by A_{q_i} and A_{q_j} are almost equivalent from the right.

<u>Proof</u>:

Let M be an augmented symetric DGSM mapping with domain R induced by the DGSM $A = \langle Q,\Sigma,\Delta,\delta,q_{in},F\rangle$. We will prove that for every pair of states q_i, q_j of A, reachable from the initial state, TƆUᗡƎЯ$_\Sigma(TR(A_{q_i} x A_{q_j}))$ is finite. To this aim, assume to the contrary that for reachable states q_1 and q_2, TƆUᗡƎЯ$_\Sigma(TR(A_{q_1} x A_{q_2}))$ is infinite.

Since M is symmetric, we can find a reversed DGSM A^R such that $\langle A,A^R\rangle$ is a symmetric pair. Let s be a prefix bound for $\langle A,A^R\rangle$ on R. We have then for all $w = u\,v \in R$,

$$\Big\| \, |A(u)| - (|A^R(u\,v)| - |A^R(v)|) \, \Big\| =$$

$$\Big\| \, |A(u)| - |A(u\,v)| + |A^R(v)| \, \Big\| =$$

(1.1) $\Big\| \, |A^R(v)| - |A_q(v)| \, \Big\| < s$ where $q = \delta_s(q_{in},u)$.

For $i \in \{1,2\}$ let w_i be the shortest word such that $\delta_s(q_{in},w_i) = q_i$. Let $p = \max\{|A(w_1)|,|A(w_2)|\}$. Choose $v \in DOM(A_{q_1} x A_{q_2})$ such that

(1.2) |TƆUᗡƎЯ$_\Sigma((A_{q_1} x A_{q_2})(v))| > 2(s+p) + 1$.

Let $A_{q_1}(v) = c_m \ldots c_1 b_k \ldots b_1$,

$A_{q_2}(v) = d_n \ldots d_1 b_k \ldots b_1$

where $k, m, n \geq 0$ and $c_i \in \Sigma$ for $1 \leq i \leq m$, $d_i \in \Sigma$ for $1 \leq i \leq n$ and $b_i \in \Sigma$ for $1 \leq i \leq k$.

By (1.2) either $m > s+p$ and ($C_1 \neq d_1$ or $n=0$)

or $n > s+p$ and ($c_1 \neq d_1$ or $n = 0$).

Without loss of generality, assume the former holds. From (1.1) it follows that $\left\| |A^R(v)| - |A_{q_1}(v)| \right\| < s$. Since $|A_{q_1}(v)| > s+p+k$, this implies that $|A^R(v)| > p+k$. Let $A^R(v) = a_\ell \ldots a_1$, where $\ell > p+k$ and $a_i \in \Sigma$ for $1 \leq i \leq \ell$. If $c_1 \neq d_1$ then either $a_{k+1} \neq c_1$ or $a_{k+1} \neq d_1$. However, $a_{k+1} \neq c_1$ implies that $A(w_1v) \neq A^R(w_1v)$ which is impossible since A and A^R are equal on R. Similarly, $A_{k+1} \neq d_1$ implies that $A(w_2v) \neq A^R(w_2v)$. Thus we must have $n = 0$. However in this case $p+k < |A^R(v)| \leq |A^R(w_2v)| = |A(w_2v)| \leq p+k$. This contradiction establishes the only if part of our proof.

For the other direction, let M be an augmented mapping induced by the DGSM $A = \{Q, \Sigma, \Delta, \delta, q_{in}, F\}$ where all states are accessable and for any pair of states q_i and q_j the mappings induced by A_{q_i} and A_{q_j} are almost equivalent from the right. We can construct a reversed DGSM for the mapping M in a manner analogous to the construction of a deterministic finite automaton to recognize the reverse of a language recognized by a given deterministic finite automaton. While reading the input backwards, we will keep track of the set of states of A which would lead to an accepting state of A from the present position, reading the reverse of what we have just read. We will also keep track of the outputs A would have had, started in any of the states in this set. Actually, we will output the common suffix of this set of "possible" outputs, retaining only the remaining initial prefixes. Using Lemma IV, it is apparent that we need only keep initial prefixes of length up to the size of the largest prefix divergence bound for any set of mappings $\{A_{q_{i_1}},\ldots,A_{q_{i_n}}\}$ for $\{q_{i_1},\ldots q_{i_n}\} \subset Q$. Thus a reversed simulation of the mapping M can be carried out using the standard "buffer" technique. □

This characterization gives us immediately:

<u>Theorem IV</u>: It is decidable whether or not an augmented DGSM mapping A is symmetric. If A is symmetric then a symmetric partner for A can be effectively constructed.

REFERENCES

[1] Engelfriet and Rozenberg, "Equality Languages and Fixed Point Languages,: Information and Control 43(1) 1979, pp. 20-49.

[2] Engelfriet and Rozenberg, "Fixed Point Languages, Equality Languages and Representation of Recursively Enumerable Languages," JACM(27), 1980, pp. 499-518.

[3] Haussler, D. and Zeiger, H.P., Very Special Languages and Representations of Recursively Enumerable Languages, to appear in Information and Control.

[4] Sipser, M. (1979) Lower Bounds on the Size of Sweeping Automata, Proc. of 11th Ann. ACM Symp. on Theory of Computing, 1979, pp. 360-364.